ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ДЛЯ РАСПОЗНАВАНИЯ СХОДСТВА ФИГУР В МЕТРИКАХ ХАУСДОРФА МЕТОДАМИ НЕГЛАДКОЙ ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ

В.И. НОРКИН, Д.А. ГУРО, М.Д. КУРШАКОВ
Национальный технический университет Украины «КПИ им. И. Сикорского»
vladimir.norkin@gmail.com

Аннотация. *Изложена информационная технология распознавания сканированных геометрических объектов с помощью методов невыпуклой негладкой оптимизации. Сначала каждый объект трансформируются в конечное множество точек евклидова пространства. Сходство между объектами устанавливается путем оптимального совмещения объектов в рамках некоторой параметрической группы преобразований и минимизации по параметрам расстояния Хаусдорфа между ними.*

Ключевые слова: *распознавание фигур, мера сходства, метрика Хаусдорфа, негладкая оптимизация, глобальная оптимизация..*

Для распознавания изображений часто используются нейронные сети, но подготовка данных и обучение сетей является трудоемкой и нетривиальной проблемой. Альтернативным подходом является распознавание изображений на основе мер сходства изображений. В данной статье рассматривается этот второй подход [1 - 3], в котором в качестве меры сходства используется расстояние Хаусдорфа (и его аналоги) между конечными множествами точек евклидова пространства.

Проблема распознавания фигур рассматривается в следующей постановке. Пусть имеется сканированное черно-белое или серое изображение в том или ином пиксельном формате. Сначала оно преобразуется в множество точек в евклидовом пространстве, первые координаты которого являются пространственными, а последующие координаты могут описывать степень яркости пикселя в градациях серого или иного цвета. Предполагается, что есть коллекция образцов геометрических фигур, которые также представлены множествами точек в двумерном или трехмерном пространстве. Распознавание состоит в поочередном оптимальном совмещении образцов с распознаваемым изображением и поиском наилучшего совпадения. Совмещающее преобразование зависит от ряда параметров, величин смещения по координатам, углов поворотов, коэффициента растяжения пространства и т.п. Например, общее аффинное преобразование двумерного евклидового пространства определяется шестью параметрами, двумерным вектором смещения и двумерной квадратной матрицей линейного преобразования. В качестве меры сходства используется метрика Хаусдорфа (или ее робастные аналоги) между двумя конечными множествами точек, а именно, точками преобразованного образца и точками изображения. Для оценки меры сходства образца и изображения решается задача нахождения параметров преобразования, которое минимизирует выбранную меру сходства. Как правило, эта оптимизационная задача является многоэкстремальной и негладкой. Для ее численного решения применяются методы негладкой глобальной оптимизации. В качестве ответа на задачу распознавания выдается тот образец, для которого получено наилучшее совпадение с изображением.

Численные эксперименты проводились с набором рукописных цифр из коллекции MNIST. Предварительно, образцы рукописных цифр, заданных в текстовом csv-формате с помощью функции fread преобразовывались в пиксельно-матричный формат, а затем в черно-белый формат. Порог отсечения неярких пикселей был выбран на уровне 240. Таким образом, каждый образец той или иной цифры представляется в виде $(m \times 2)$-матрицы M, где m - индивидуальное число точек представления цифры на плоскости, а столбцы матрицы содержат горизонтальные и вертикальные координаты этих точек. Нелинейное преобразование совмещения цифр включало смещение одной цифры в начало координат, поворот, растяжение по осям и перемещение к другой цифре. Пусть матрицы M^1 и M^2 задают представления двух черных цифр-фигур X^1 и X^2 на плоскости с числами точек m_1 и m_2. Обозначим $x_i^1 = \left(x_{i1}^1, x_{i2}^1\right)$, $i = 1,...,m_1$, и $x_i^2 = \left(x_{i1}^2, x_{i2}^2\right)$, $i = 1,...,m_2$, строки этих матриц.

Нелинейное преобразование цифры-фигуры X^1 в фигуру $Y^1=\left\{y_i^1=\left(y_{i1}^1,y_{i2}^1\right),i=1,...,m_1\right\}$ для совмещения с X^2 задается пятью переменными параметрами $p=\left(p_1,...,p_6\right)$:

$$y_i^1=\left(y_{i1}^1,y_{i2}^1\right)=(x_{i1}^1-\overline{x}_{i1}^1,x_{i2}^1-\overline{x}_{i2}^1)\begin{pmatrix}p_1 & p_2\\ p_3 & p_4\end{pmatrix}+\left(p_5,p_6\right),$$

где $\overline{x}_{i1}^1=\left(1/m_1\right)\sum_{i=1}^{m_1}x_{i1}^1$, $\overline{x}_{i1}^2=\left(1/m_1\right)\sum_{i=1}^{m_1}x_{i1}^2$, а преобразование цифры-фигуры X^2 в фигуру $Y^2=\left\{y_i^2=\left(y_{i1}^2,y_{i2}^2\right):i=1,...,m_2\right\}$ для совмещения с X^1 задается переменными параметрами $p'=\left(p_1',...,p_6'\right)$:

$$y_i^2=\left(y_{i1}^2,y_{i2}^2\right)=(x_{i1}^2-\overline{x}_{i1}^2,x_{i2}^2-\overline{x}_{i2}^2)\begin{pmatrix}p_1' & p_2'\\ p_3' & p_4'\end{pmatrix}+\left(p_5',p_6'\right),$$

где $\overline{x}_{i1}^2=\left(1/m_2\right)\sum_{i=1}^{m_2}x_{i1}^2$, $\overline{x}_{i2}^2=\left(1/m_2\right)\sum_{i=1}^{m_2}x_{i2}^2$. Оптимальные значения параметров преобразований находятся как решения следующих оптимизационных задач:

$$\Delta_H(Y^1(p),X^2)\to\min_{p\in\Pi},\qquad \Delta_H(Y^2(p'),X^1)\to\min_{p'\in\Pi}.$$

Эти задачи являются задачами негладкой глобальной оптимизации на множестве

$$\Pi=\left\{p\in\mathrm{R}^6:p_1\in[0,2];-1\le p_2,p_3\le 1;p_4\in[0,2];p_5,p_6\in[0,1]\right\}.$$

В данной работе они решались генетическим алгоритмом и методом последовательного сглаживания [4]. В качестве меры сходства цифр-фигур X^1 и X^2 используется величина

$$D_H(X^1,X^2)=\max\left\{\min_{p\in\Pi}\Delta_H(Y^1(p),X^2),\min_{p\in\Pi}\Delta_H(Y^2(p),X^1)\right\}.$$

Вместо отклонения Хаусдорфа Δ_H могут использоваться его разнообразные робастные аналоги. Решение о совпадении цифр-фигур X^1 и X^2 принимается на основании значения меры сходства $D_H(X^1,X^2)$. Например, решение задачи классификации по методу ближайшего соседа, т.е. решение о принадлежности цифры-фигуры

X^* к классу фигур F принималось, если расстояние фигуры X^* до класса фигур F, $\min_{X \in F} D_H(X^*, X)$, было минимальным по сравнению с расстояниями до других классов. Пример оптимального преобразования и совмещения цифр-фигур показан на рис. 1, 2.

РИС. 1. Два варианта написания цифры 3 в черно-белом пиксельном формате.

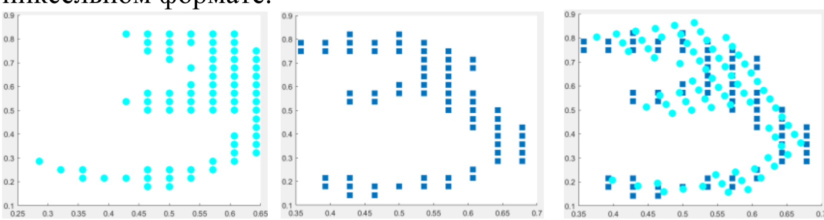

РИС. 2. Представление цифр в векторном формате и результат их трансформации и совмещения.

Литература.

1. Rucklidge W. Efficient visual recognition using the Hausdorff distance. *Lecture notes in computer science ; Vol. 1173*. Berlin; Heidelberg; New York: Springer, 1996. 178 p.
2. Ушаков В.Н., Лебедев П.Д. Итерационные методы минимизации хаусдорфова расстояния между подвижными многоугольниками. *Вестн. Удмуртск. ун-та. Матем. Мех. Компьют. науки*. 2017. **27,** Вып. 1. С. 86-97. https://doi.org/10.20537/vm170108
3. Kumar K.S., Manigandan T., Chitra D., Murali L. Object recognition using Hausdorff distance for multimedia applications. *Multimed Tools Appl*. 2020. **79**. P.4099-4114. https://doi.org/10.1007/s11042-019-07774-z
4. Norkin V.I. A stochastic smoothing method for nonsmooth global optimization. *Cybernetics and Computer Technologies*. 2020, Issue 1, P. 5-14. https://doi.org/10.34229/2707-451X.20.1.1